Introduction to Computational Game Theory

CMPT 882

Simon Fraser University

Oliver Schulte

Rational Preferences

(start with powerpoint)

Weak Preferences

- Let O be a set of options among which an agent A is choosing. The options can be stocks, ice cream flavours, potential spouses, jobs, cities, universities, anything. For simplicity we'll assume that there are only finitely many options.

- We assume that the agent has a **weak preference relation** $\geq$ among the options in O.

There are several ways to interpret this relation, for example:

1. $x \geq y$ means that the agent prefers x *at least as* much as y.

2. $x \geq y$ means that the agent likes x strictly better than y **or** is indifferent between x and y.

3. $x \geq y$ means that the agent would *trade* y for x.

Note that the first two interpretations refer to the agent's inner or *mental properties*. The third refers to the agent's *behaviour*.
The **revealed preference** thesis states that the agent's choices correspond to her preferences:

Choice Reveals Preference

Strict Preferences

Sometimes we write $x > y$, read:
the agent **strictly prefers** x to y.

This means that

1. The agent prefers x more than y.
2. The agent likes x better than y.
3. The agent would trade y for x but not the other way around.

As before, we assume that choice reveals preference.

Defining Preference Relations

Some preference relations are rankings of the available options. (We'll come back to this.) Rankings of options can be **represented** by assigning a number to each option, such that an option is preferred to another just in case its number is greater. (Think of numbers of votes or ridings in an election.)

Example

Suppose that Jane ranks TV shows as follows: Buffy the Vampire Slayer is the best, then Friends, then the Simpsons, finally Angel. Let ≥ be Jane's preference ordering among these shows. Then we can define ≥ as follows.

score(Buffy) = 4
score(Friends) = 3
score(Simpsons) = 2
score(Angel) = 1

Note that for all options $x \geq y$ iff score(x) > score(y).

We could also define ≥ as follows.

score2(Buffy) = 1,000
score2(Friends) = 500
score2(Simpsons) = -1000
score2(Angel) = -10,000

Properties of the Weak Preference Relation

We postulate that a *rational* weak preference relation $\geq$ has the following properties.

1. **Totality**: For all options x,y either $x \geq y$ or $y \geq x$.

(we allow that both $x \geq y$ and $y \geq x$, in which case the agent is indifferent between x and y.)

2. **Reflexivity:** For all options x, it is the case that $x \geq x$.

3. **Transitivity:** If $x \geq y$ and $y \geq z$, then $x \geq z$ for all options x,y,z.
Totality says that the agent can always decide between any two options, in the sense that she can always decide whether she would trade one for another.

- These are very general principles that apply to **any** choice situation, regardless of whether it involves uncertainty, complex options etc.
- Reflexivity and transitivity seem natural enough. Moreover, there is a powerful argument for why an agent should satisfy these requirements: the **money pump argument**.

The Money Pump Argument

Let O be a set of options among which an agent A has a weak preference relation $\geq$. Suppose that $\geq$ is not transitive. Then there are options x, y, z such that

$x \geq y \geq z$ but it is not the case that $x \geq z$. So we have that

$x \geq y \geq z > x$.

Exercise

Suppose that Fred has intransitive preferences of the kind described above, such that $x \geq y \geq z > x$. Assume that whenever Fred weakly prefers an option a to an option b, Fred is willing to trade b for a. Assume further that whenver Fred strictly prefers an option a to an option b, Fred is willing to pay 1\$ and b to get a. Show how you can turn Fred into a money pump, that is how you can get him to give you any amount of money for nothing.

Rational Preferences

Definition

A weak preference relation ≥ is **rational** iff ≥ is total, reflexive and transitive.

Theorem

Given a finite set of options, there is a score function that represents a weak preference relation ≥ if and only if ≥ is rational.

In light of this theorem, we can go back and forth between score functions and preference relations to represent rational preferences. We will refer to score functions as **payoff functions** or as **utility functions**, using the notation u(o)**.**